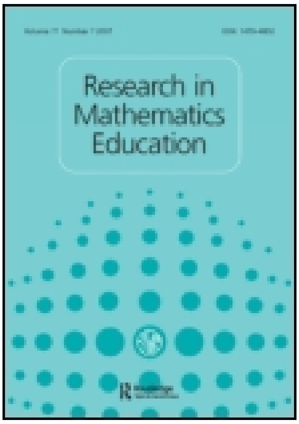

Research in Mathematics Education

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rrme20

On the different ways that mathematicians use diagrams in proof construction

Aron Samkoff [a] , Yvonne Lai [b] & Keith Weber [a]

[a] Rutgers University, USA

[b] University of Michigan, USA

Available online: 22 Feb 2012

To cite this article: Aron Samkoff, Yvonne Lai & Keith Weber (2012): On the different ways that mathematicians use diagrams in proof construction, Research in Mathematics Education, 14:1, 49-67

To link to this article: http://dx.doi.org/10.1080/14794802.2012.657438

Research in Mathematics Education
Vol. 14, No. 1, March 2012, 49–67

On the different ways that mathematicians use diagrams in proof construction

Aron Samkoff[a]*, Yvonne Lai[b] and Keith Weber[a]

[a]*Rutgers University, USA;* [b]*University of Michigan, USA*

The processes by which individuals can construct proofs based on visual arguments are poorly understood. We investigated this issue by presenting eight mathematicians with a task that invited the construction of a diagram, and examined how they used this diagram to produce a formal proof. The main findings were that participants varied in the extent of their diagram usage, it was not trivial for participants to translate an intuitive argument into a formal proof, and participants' reasons for using diagrams included noticing mathematical properties, verifying logical deductions, representing ideas or assertions, and suggesting proof approaches.

Keywords: mathematicians; diagrams; proof construction

Introduction

Student difficulty with proof

A primary goal of advanced undergraduate mathematics courses is to improve students' abilities to construct proofs. Yet numerous studies have documented undergraduates' frequent inability to construct mathematical proofs (e.g. Alcock and Weber 2010a; Hart 1994; Moore 1994; Recio and Godino 2001; Weber 2001; Weber and Alcock 2004). Research in this area has generally focused on particular difficulties that undergraduates face when writing proofs, such as having a poor understanding of advanced mathematical concepts (Hart 1994; Moore 1994), lacking proving strategies (Weber 2001), and not knowing where to begin when asked to write a proof (Moore 1994). However, research on how students can, or should, engage successfully in proof construction has been sparse. In this paper, we examine in detail one specific suggestion from the mathematics education literature – using a diagram as a basis for constructing a formal proof.

Diagrams in mathematics

Diagrams are viewed by mathematicians and mathematics educators alike as an integral component of doing and understanding mathematics (e.g. Hadamard 1945; Stylianou 2002). A substantial benefit that diagrams afford is that they allow the problem solver access to view, compare, and integrate simultaneous pieces of information with less cognitive effort than when the same information is presented symbolically and sequentially (Dreyfus 1991) – while making more transparent

*Corresponding author. Email: samkoff@gmail.com

ISSN 1479-4802 print/ISSN 1754-0178 online

http://dx.doi.org/10.1080/14794802.2012.657438
http://www.tandfonline.com

certain properties of mathematical concepts that would ordinarily be difficult to discern with non-visual representations of this concept (e.g. Piez and Voxman 1997). Further, diagrams can be used to provide novel and more accessible explanations for mathematical phenomena, or highlight aesthetics that are less accessible through symbols and logic (e.g. Hersh 1993), as is illustrated by Nelsen's (1993, 2000) *Proofs without words*. Mathematicians' self-reports reveal that diagrams can play a significant role in their mathematical work (e.g. Dreyfus 1991; Hadamard 1945). Stylianou's (2002) investigation into mathematicians' problem solving revealed four purposes for diagrams, which she terms as inferring consequences, elaborating upon mathematical ideas, creating sub-goals, and encouraging metacognitive reasoning.

Drawing diagrams is commonly cited as a heuristic for mathematical problem solving that students should engage in (e.g. Polya 1957; NCTM 2000; Schoenfeld 1985). However, despite the promise of this recommendation, the literature suggests that at least several pedagogical issues must be addressed. First, researchers have noted that students are often reluctant to use diagrams, even for problems where their use might be highly productive (e.g. Dreyfus 1991; Piez and Voxman 1997; Stylianou and Silver 2004). Second, Presmeg (1986) found little correlation between high school students' propensity to visualise and their mathematical performance. Indeed, the strongest students in Presmeg's sample rarely used diagrams. Finally, Schoenfeld (1985) demonstrated that the process of implementing this type of heuristic is surprisingly complex; he argued that more explicit instruction on using diagrams (e.g. what to read from diagrams) might be necessary for students to employ this heuristic effectively.

Diagrams and formal proofs

Instructors of advanced undergraduate courses often ask their students to produce formal proofs. These proofs are expected to begin with definitions, axioms, and appropriate assumptions, and proceed deductively to reach a desired conclusion, often while employing logical syntax. While informal representations of the involved concepts, including diagrams, may be included with the presentation of the proof, their role is expected to be ancillary, serving as a comprehension aid for the formal argument. The inferences within the proof are expected to be based on deductive logic, not the appearance of the diagram.

Both mathematicians and mathematics educators emphasise that although the product of proving is a formal argument, the process of proving can be, and frequently is, based on informal argumentation – often involving visual reasoning (e.g. Alcock 2010; Boero 2007; Burton 2004; Raman 2003; Thurston 1994; Weber and Alcock 2004). Consequently, some researchers have suggested that undergraduates be encouraged to base the proofs that they construct on informal reasoning (Boero 2007)—in particular by utilising diagrammatic reasoning (Alcock 2010; Gibson 1998; Raman 2003; Weber and Alcock 2004).

The literature supports this suggestion both theoretically and empirically. Theoretically, researchers have noted benefits offered by diagrammatic reasoning in formal mathematical settings. Diagrams can reify formal mathematical concepts to make them meaningful for students (Alcock and Simpson 2004; Gibson 1998), allow students to draw useful inferences quickly (Alcock and Simpson 2004;

Gibson 1998), and help students overcome the impasses they reach when writing a proof (Gibson 1998; Weber and Alcock 2004). Empirically, there are case studies documenting undergraduates' success in producing proofs based on diagrams (Alcock and Weber 2010a; Gibson 1998), demonstrating that a diagrammatic approach to writing proofs can be useful for some students in some situations. Furthermore, there is strong evidence that some mathematicians frequently base their proofs on diagrams (e.g. Burton 2004; Hadamard 1945), suggesting that diagram usage is representative of many mathematicians' practice.

Limitations to the effectiveness of basing proofs on diagrams

Although there is good reason to believe that undergraduates should base some proofs on diagrams, many researchers have delineated theoretical difficulties that might prevent students from successfully basing a proof on a diagram. For instance, Duval (2007) noted that the structure of a visual argument (or any informal argument) differs significantly from the structure of a formal proof; he cautions researchers not to underestimate the cognitive complexity of determining the statuses of assertions within a proof (i.e. is the assertion an assumption, an axiom, an accepted fact, or a deduction?) and organising the chain of deductions appropriately. This can be particularly difficult when basing a proof on a diagram, as each assertion may appear equally obvious, making it hard for students to distinguish what can be assumed and what needs to be shown (cf. Alcock and Simpson 2004).

There are theoretical difficulties beyond cognitive complexity with using diagrams effectively in proof construction. For instance, Alcock and Weber (2010a, 2010b) argue that students often have inaccurate representations of mathematical concepts, or fail to see connections between their visual representations of a mathematical concept and the concept's formal definition, making it difficult or impossible to base formal proofs on those diagrams.

Moreover, although some published case studies in mathematics education illustrate how some undergraduates are able to use insight gleaned from diagrams to construct proofs, other case studies illustrate how undergraduates are often not able to do so. For instance, Alcock and Weber (2010b) asked eleven mathematics majors to prove that a real-valued increasing function does not have a global maximum. The four students who drew diagrams of a generic increasing function each instantly gained conviction that the statement to be proven was true, but none could see how to begin writing a proof. Finally, several small-scale studies have found little or no correlation between undergraduates' propensity to use diagrams and their success in proof-writing in advanced mathematics (e.g. Alcock and Simpson 2004, 2005). These findings suggest research is needed on how, specifically, students might use diagrams in their proof construction.

Research questions

Reflecting on her own teaching, Alcock (2010) writes:

> Diagrams can provide insight, but it is not always easy for students to make detailed links between what is in the diagram and what is in a formal proof. This means that the

> step between seeing that a result must be true and proving it can seem insurmountable. Through my small-class teaching, I have also learned that students often find it difficult to draw a diagram based on verbal and algebraic information. As a result, I now spend more time walking students through the process of drawing diagrams (232–33).

We share Alcock's goal of making explicit the process that connects the construction of formal proofs and the gaining of conviction from diagrams. The study reported in this paper contributes to this goal by examining how eight mathematicians use diagrams in their own proof-writing. Two questions drove our analysis:

- To what extent did each of these mathematicians base their proofs on the diagrams?
- How, and for what purposes, did each of these mathematicians use diagrams to assist their proof writing?

The analysis in this paper is qualitative and interpretive. Consequently, we will not attempt to generalise our claims to the general population of mathematicians. Rather, the goal of this analysis is to provide illustrations of how individual mathematicians reasoned about this task. This allows us to illustrate the variety of diagram usage mathematicians may exhibit while completing the same task and the difficulties that some mathematicians had with translating a diagram into a proof, while showing the varied ways in which they used diagrams. However, we refrain from making general claims of the form that the behaviours we observed would be common among mathematicians, and from speculating about what percentage of mathematicians would exhibit the types of reasoning we report. Such questions would need to be addressed in future research, employing larger sample sizes.

Methods

Participants

The second author, a mathematics research post-doctoral fellow during the data collection period, emailed five personal contacts at two private universities and two large state schools, in four different states in the USA, about the study; she then invited these contacts, and colleagues recommended by them, to participate. Among the 12 mathematicians invited, 10 agreed to participate. We strove to obtain participants who were articulate and reflective, and who were known to be promising or successful research mathematicians. However, we could only use data from eight participants in this study – one mathematician produced something resembling lecture notes for himself instead of a full written proof, and another mathematician's data was inaccessible due to a video-recording error. Our report is based on analysis of the remaining eight mathematicians' work. More information about the participants' experience as mathematicians is contained in Table 1. Their experience varied, ranging from 1 year to 30 years, with four of them having at least 5 years' experience working in mathematics departments. All these mathematicians currently work, either in post-doctoral positions or in tenure-track positions, at a research institution in one of the top 25 graduate programs in mathematics in the United States. They are employed at a combination of three large state, and two private, universities in three

Table 1. A summary of the mathematicians' experience.

Participant	Time since acquiring Ph.D. (in years, approximate)	Number of papers refereed (approximate)	Position (at the time of participating in the study)
BT	30	40	Professor
KT	25	180	Professor
CY	20	40	Professor
IR	10	5	Assistant professor
KZ	2	4	Assistant professor
TL	2	3	Post-doctoral fellow
PV	1	4	Post-doctoral fellow
FR	1	2	Post-doctoral fellow

different regions of the USA. In this report, we use pseudonym initials to refer to the participants, and refer to all with the masculine pronoun to further protect their identities. The pseudonym initials are not the initials of their real names.

Materials

Participants were asked to complete the Sine Task given below:

> Show that restrictions of the sine function to intervals of length greater than π cannot be injective.

The Sine Task features the following characteristics. First, the claim is based on relatively elementary mathematics – the sine function and injectivity are part of most secondary mathematics curricula. Thus, the meaning of this claim would be accessible to second and third year undergraduate mathematics majors. Secondly, proving this claim invites the use of a diagram; we believed it would be natural for mathematicians to graph the sine function on reading the claim. Thirdly, we believed that it would be fairly easy to accept the claim as true after viewing the graph of a sine function, based on pilot data that was collected. Hence the participants would not need to spend a long time verifying that the claim was correct; rather, most of the work in proving this claim would involve creating an argument to establish it. Fourthly, despite the relative simplicity of the mathematical concepts involved and the apparent truth of the claim, we believed that writing a full proof would not be trivial. Our results confirm that these assumptions about the Sine Task were accurate. As a consequence of the above characteristics, observing mathematicians' performance on this task has the potential to give insight into the processes that mathematicians use to construct a proof based on a diagram, in a mathematical domain accessible to undergraduate mathematics majors.

Procedure

Participants met individually with the second author for a one-hour interview. All interviews were video-taped. Participants were handed the task described above, and were told to think aloud while writing a proof of this claim that would be appropriate for a textbook for sophomore – or junior-level mathematics majors. After producing a proof, participants were invited by the interviewer to revise the proof, using a different coloured pen. Participants were given unlimited time to complete this task. They were also told to use rough paper as needed, and these papers were also collected and analysed.

Analysis

The participating mathematicians collectively made a total of 321 mathematical statements. We coded these statements in terms of their type and source. A statement was coded as an *assertion* if it was a mathematical statement that could be validated as true or false. An example of an assertion is "$\sin(x+\pi) = -\sin(x)$". A statement was coded as a *proof approach* if it suggested a way to proceed with the proof considered by the participant, such as a suggestion to use a proof by cases or to employ techniques from calculus. A statement was coded as an *evaluation* if the participant judged the validity of a previously stated assertion (e.g. "$\sin(x+\pi) = \sin(x)$ is false"), the utility of a previously stated assertion, or the viability of a previously stated proof approach. We were able to code each mathematical statement into one of these three categories.

We next considered the source of each of the statements that participants made. We attributed statements to three sources. A *deduction* was an assertion that was a logical consequence of previous statements. Deductions included instances when a participant specifically referred to previous statements that he or she had made, indicated they were making a new deduction by using a connector such as "hence" or "consequently", or uttered a statement that was an immediate mathematical consequence of the statement preceding it. A statement was coded as an *inference from a diagram* if the participant cited, either verbally or with gesture, a graph of the sine function, or another diagram that he or she had produced as a basis for making that statement. A statement was coded as *recall or unknown* if the participant neither referenced a graph or diagram that he or she had constructed, nor cited previous statements as a basis for making the new statement. Finally, a statement was coded as *other* if it did not fit neatly into one of these categories. For example, reiterations of previously made statements were coded under this category. In all, 47 statements were coded as deductions, 32 as inferences from a diagram, 206 as recall/unknown, and 36 were coded as other.

When a statement was coded as a deduction, it was also noted from which statement(s) it was deduced. Thus, it was possible to determine whether or not statements of the final proof were initially generated using a diagram.

For each inference made from a diagram, we used an open coding scheme in the style of Strauss and Corbin (1990) to classify the purpose that the diagram served. We first wrote a short description of each inference. Similar inferences were grouped together, and given preliminary category names and definitions. New inferences were placed into existing categories when appropriate, but also used to create new categories, or modify the names or definitions of existing categories. This process continued until a set of categories was formed, grounded to fit the available data. Descriptions and illustrations of each category are provided in Section 4.4.

Finally, we coded the correctness of the participants' final proofs using the coding scheme of Malone et al. (1980) that was used by Hart (1994) and Weber (2006) in their studies evaluating the correctness of students' proofs. Each proof was coded as completely correct, correct except for minor and insignificant errors, incorrect with significant errors but substantial progress, or incorrect with no significant progress. Typed versions of two completely correct proofs are presented in the Appendix.

Results

A summary of the participants' behaviour is presented in Table 2.

Table 2 reveals that the participants varied in their diagram usage. In this section, we first present descriptions of four participants chosen to illustrate the differing extents that participants used diagrams. We chose two participants who used diagrams extensively (KZ and TL), one whose use of diagrams was limited (FR), and one who did not use diagrams at all (KT). In section 4, we use these descriptions as a basis to address our research questions.

KZ

Representing the situation and noticing a property

After reading the instructions, KZ immediately drew a graph of one period of the sine function and stated:

> So now I want to think about for which x and y I have $\sin(x) = \sin(y)$, so if I look at the graph, I can see that this is $\pi/2$, this is $3\pi/2$ [drawing vertical lines at these x values on the graph], so from here I can see that if I take ... sine of π [sic] plus theta is equal to sine of π [sic] minus theta.

Table 2. A summary of the mathematicians' proofs. (Key: NP = Noticing properties/generating conjectures, RI = Representing/instantiating ideas or assertions, ET = Estimating the truth of an assertion, SPA = Suggesting a proof approach).

Mathematician	Drew a diagram?	Quality of proof	Time on task (minutes)	Number of inferences from the diagram	Purposes served by diagram
TL	Yes	Mostly correct	30	10	NP, RI, SPA
CY	Yes	Mostly correct	11	2	NP
PV	Yes	Correct	9	7	NP, ET
FR	Yes	Incorrect with substantial progress	18	2	NP, ET
IR	Yes	Correct	11	3	NP, ET
BT	Yes	Needed substantial hints	21	2	RI, SPA
KZ	Yes	Correct	33	6	NP, RI, ET, SPA
KT	No	Correct	14	0	

This excerpt shows how KZ deliberately used the graph to search for two values which have the same sine value, which leads him to notice a general property of the sine function. He next writes $\sin(\pi + \theta) = \sin(\pi - \theta)$, soon changing this to the correct equation

$$\sin(\pi/2 + \theta) = \sin(\pi/2 - \theta) \tag{1}$$

Writing and informally justifying a key formula

KZ attempted to prove equation (1) using the identity

$$\sin(\theta) = \cos(\pi/2 - \theta) \tag{2}$$

He justified (2) by noting this "can be seen if you think of sine from the triangle," and he drew a right-angled triangle with one acute angle labeled θ and the other, $\pi/2–\theta$. He then wrote an equation analogous to (1) for the symmetry of the sine function at $x = 3\pi/2$:

$$\sin(3\pi/2 + \theta) = \sin(3\pi/2 - \theta) \tag{3}$$

He declared that equations (1) and (3) were important to the proof, and began to focus on how to use these identities in the proof.

Building an informal proof

KZ asserted that every interval of length greater than π must contain an element of the form $k\pi/2$, where k is an odd integer, as well as an interval around this point. He sketched a real number line, labeled some odd multiples of $\pi/2$ on it, and argued that since these points occur every π units, every interval of length greater than π must include one of these points, as well as an interval around such points. This is illustrated in the following excerpt:

> So this interval [*a*, *b*] must contain an element of the form $k\pi/2$ because [begins drawing a number line] we're talking about $\pi/2$, $3\pi/2$, $5\pi/2$ and so on, the distance between those is π [labels these points on the number line as he states them]. So if you take any interval of length strictly bigger than π, it must contain at least one of them. [draws the interval of length greater than π on the number line] And moreover, it must contain some interval close to them too, both on the left and on the right, and you can choose a θ close enough there.

Throughout this excerpt, KZ is representing the ideas that he says aloud on his diagram of the number line. Apparently satisfied with this part of the argument, KZ then used (2) to deduce (1), and declared himself ready to write a final draft of the proof. However, as described below, KZ struggled to establish (1) from more elementary principles in a rigorous way.

Attempting to write a proof

KZ began his final proof by rewriting the claim in a more formal way. To prove this version of the claim, he decided to prove a more general statement:

$$\sin(k\pi/2 + \theta) = \sin(k\pi/2 - \theta) \text{ for all integers } k \tag{4}$$

He employed an argument similar to that used in the first draft, based on the identity (2), but decided to try to prove this latter fact. He attempted to do this via the cosine angle addition formula, but misremembered the formula and deduced:

$$\cos(\pi/2 - \theta) = -\sin(\theta) \tag{5}$$

KZ was disturbed by the contradiction between (2) and (5), so he again tried to deduce (2) from the cosine angle addition formula, but made the same mistake.

Resolving the contradiction

To resolve this contradiction, KZ drew two graphs – one of the sine function and one of the cosine function – and, while looking at the graphs said, "plugging in something for sine is the same as plugging in something minus $\pi/2$ for cosine. It should be right; maybe I got my identity wrong?" He then evaluated the incorrect identity at $\theta = 0$ and $\pi/2$, noticing the identity fails for the latter value. He immediately recognised the error in the identity and corrected the mistaken sign, then continued with the proof.

It is interesting to note how KZ became convinced that (2) was true. Some researchers claim that mathematicians become fully convinced in an assertion only via deductive reasoning and, moreover, should find deductive arguments that they produce to be completely convincing (e.g. Harel and Sowder 1998). However, when KZ deduced (5), he did not accept (5) as true and abandon (2), since (2) seemed consistent with a right-angled triangle he drew. To distinguish between (2) and (5), he then turned to graphs, and ultimately to examples. This illustrates how deductive reasoning alone is not always the final arbiter of whether an assertion is correct for all mathematicians, but may be coordinated with diagrammatic, graphical, and empirical reasoning to decide on the truth of an assertion. In this case, conviction came prior to formal proof (cf. Thurston 1994).

Finishing the proof

Having established the identity, he argued that for any odd integer k, $\sin(k\pi/2 + \theta) = \sin(k\pi/2 - \theta)$ by appealing to the evenness of cosine function. He finished his final draft by arguing that any interval of length greater than π must contain an element of the form $k\pi/2$, for odd k, as well as a neighborhood around such an element. KZ's final proof was coded as "Correct".

TL

Choosing a proof approach

After reading the task, TL drew a graph of two periods of the sine function, and labeled several x-values on the graph (see Figure 1). He gestured and drew on this graph as he considered possible approaches to prove the claim. After briefly considering using techniques from calculus to approach the problem, TL stated:

> So I already have the picture of it [referring to Figure 1], well let's see, so it's a periodic function. So, whatever interval that you were looking at – if you take some interval of length strictly greater than π, you can always translate it backwards so that it's around the origin, because sine is periodic. OK, so here's a period of the sine function from here [draws the two dashed lines on the sine graph in Figure 1]. OK, and we want to know about intervals of length strictly greater than π. So if we take any interval at all, I can move it back to this region, and it's necessarily going to contain the origin because length of the interval is strictly greater than π [draws an interval containing the origin on the graph]. I guess this is an easier proof than taking derivatives, but maybe it's not quite a proof yet. I have to think about it a little more carefully. OK. So we take any interval of length strictly greater than π and move it back into this fundamental region here . . . So the question is I have to make a precise way of saying what it is that I'm talking about, but I can see how the argument is forming.

In this excerpt, TL lays out how his proof will proceed. He will use the periodicity of the sine function to "translate back" an arbitrary interval so as to contain the origin.

Next, TL drew an interval containing the origin on the graph, and considered somehow taking advantage of the fact that the sine function is odd, but he later abandoned this idea. He then remarked that, if the interval contains $\pi/2$ or $-\pi/2$ in its interior, the function cannot be injective on this interval "sort of obviously", as illustrated by the following excerpt:

> [While gesturing to graph in Figure 1] OK, so clearly, let's see. If the interval strictly contains this point, π/2 or minus π/2, then you can see from these areas . . . So we take this interval and we translate it back, so if that interval contains either one of these, then it's not going to be injective, sort of obviously.

TL went on to mention Rolle's Theorem, and the fact that the derivative of the sine function at $\pi/2$ and $-\pi/2$ is zero, but delayed a more thorough investigation as to why this statement is true. Finally, he asserted that an arbitrary interval can be translated back so as to contain $\pi/2$ or $-\pi/2$ in its interior. Seemingly satisfied with the overarching plan of his proof, TL began writing a more complete version of the proof.

Establishing a relationship between injectivity and local maxima of a continuous function

Using techniques from calculus, TL quickly established that sine has a local maximum at $x = \pi/2$. He then claimed that a function cannot be injective on a neighborhood containing a local maximum, and justified this verbally as following "immediately from the definition of what it means to be a local maximum." To justify this claim, TL wrote the definitions for injectivity and a local maximum, and drew several sketches of

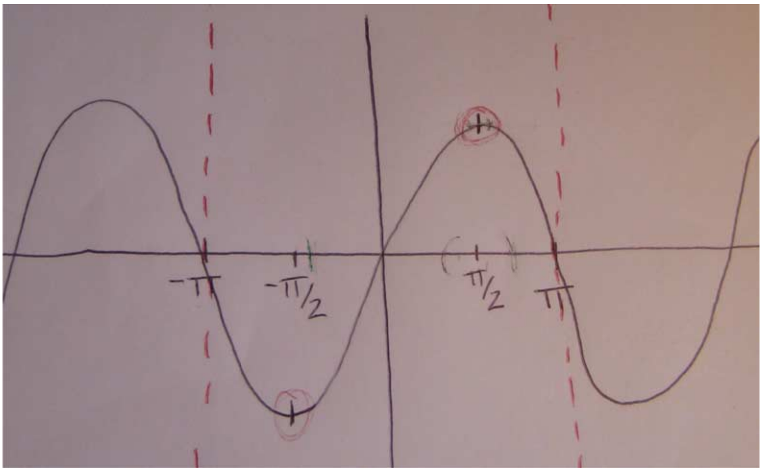

Figure 1. TL's graph of the sine function.

a continuous function with a local maximum (see Figure 2). After drawing the first of these sketches, he remained silent for some time, but finally uttered:

> Let me think for a minute. [draws middle diagram in Figure 2] OK, so let me see. So this is going to be something like Rolle's Theorem or the Intermediate Value Theorem . . . is basically the tool that I have to use, but I just have to figure out the right way of saying it. I don't think I can just appeal to just a theorem to do it . . .

After four minutes of thought, TL developed an argument for this sub-claim by picking points, *b,* and *c*, to the left and right, respectively, of *a,* the *x*-value of

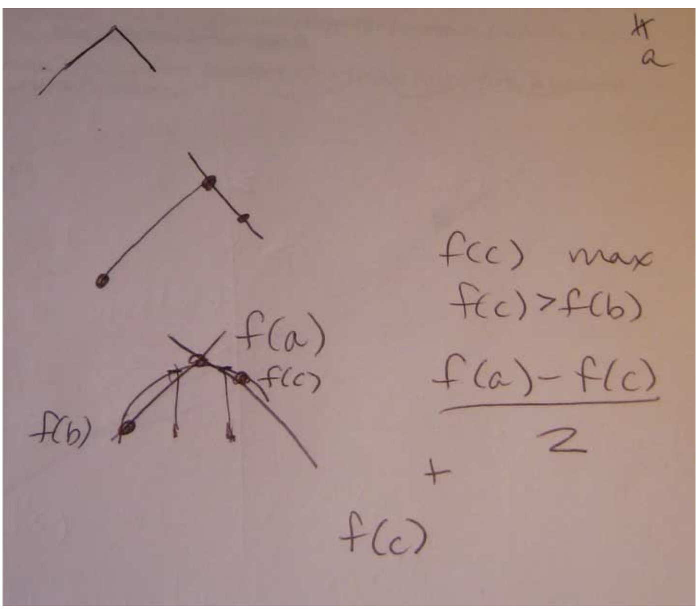

Figure 2. TL's continuous functions with local maxima.

the local maximum, and applying the Intermediate Value Theorem to show that the function fails to be injective, by showing that the average of $f(a)$ and max $\{f(b), f(c)\}$ was mapped to twice in the interval. (We note that it would also have sufficed to show max$\{f(b), f(c)\}$ is mapped to twice).

Writing a final proof

TL's final proof argued that it was sufficient to prove the claim on the interval $[-\pi, \pi]$. He showed that if the interval contains $x = \pi/2$, sin x would not be injective on that interval, since $x = \pi/2$ is a local maximum. However, the proof did not consider intervals that did not contain $x = \pi/2$ but instead contained the local minimum $x = -\pi/2$. As the argument for this other possibility is essentially the same as what TL produced, we coded the proof as "Mostly correct".

FR

Rejecting an initial assumption

FR initially conjectured that $\sin(x) = \sin(x+\pi)$. He then drew a graph of the sine function, and quickly rejected this conjecture, as the following excerpt illustrates:

> Sine of x and sine of x plus π, wait, are those equal? [draws a graph of one period of the sine function] I'm trying to remember the period of π [sic] ... the period of π[sic] is 2π, so it would be easy to show the sine function was not injective on intervals of length 2π. What about intervals of length greater than π ... Let's see ... so, this is not true that they're equal. False.

After this point, there was no evidence of any further diagram usage by FR.

Writing a flawed proof of the theorem

FR proceeded to produce a proof in which Rolle's Theorem is used to deduce that $\sin(x)$ cannot be injective on an interval where its derivative attains a zero. However, Rolle's Theorem asserts that given real numbers a, b with $a < b$, if a continuous differentiable function f satisfies $f(a) = f(b)$, then there exists a real number c such that $a < c < b$ and $f'(c) = 0$. FR was actually applying the false converse of Rolle's Theorem, assuming that if there were a c such that $f'(c) = 0$, and f were defined on an interval around c, there must be distinct points a and b such that $f(a) = f(b)$. (Note for instance, this claim is false for $f(x) = x^3$ due to an inflection point, although the claim is true in the trigonometric case relevant to FR). For this reason, FR's proof was coded as "Incorrect with substantial progress".

KT

KT's work was unusual in that during the writing of his proof, he drew no diagrams, and he worked in a rapid linear fashion. He produced two proofs in quick succession, deeming the second one an improvement over the first. A copy of the first proof may be found in the Appendix (proof 2). Since both proofs were correct, KT's response was coded as "Correct".

General findings about participants' use of diagrams

Overall use of diagrams

We predicted that the Sine Task would invite the use of a diagram, and that this diagram would give participants confidence that the claim in this task was correct. The participants' behaviour verified these assumptions. Seven out of eight drew a graph of the sine function shortly after reading the problem statement. Six participants indicated that they were convinced t the claim was true either prior to, or immediately after, drawing the graph.

Three participants drew diagrams aside from the graph of the sine function. KZ drew a right-angled triangle while attempting to verify that $\sin(\theta) = \cos(\pi/2 - \theta)$; later, he drew a number line to help him justify that every interval $[a, b]$ of length strictly greater than π contained a point of the form $k\pi/2$ for some odd integer k. TL drew a portion of the graph of a function containing a local maximum to prove functions cannot be injective on intervals containing local maxima. When BT reached an impasse in writing the proof and did not know how to proceed, he drew a unit circle, and made two inferences from this diagram.

Variance in the extent of participants' diagram usage

Our descriptions in section 3, as well as Table 2, illustrate the different extents to which participants used diagrams. At one extreme, diagrams pervaded nearly every aspect of KZ's proof-writing. He used a diagram to represent the claim being proved (see 3.1.1), generate the key formula for his proof (3.1.2), informally justify this formula (3.1.2), and resolve a contradiction by locating a logical error (3.1.5). TL (section 3.2) behaved similarly to KZ, and IR and PV used diagrams to draw inferences that were ultimately connected to their final proofs.

At the other extreme was KT (section 3.4), who did not draw a diagram at all during his proof construction. FR only used a diagram to reject an incorrect conjecture about the period of the sine function, and to identify the correct period (section 3.3.1) but for nothing else. Similarly, BT only considered a diagram when he reached an impasse in his proofs. The inferences he generated from this diagram were not useful in producing a proof.

In summary, the question of the extent to which participants' used diagrams did not have a uniform answer across participants, and highlights the heterogeneity in how mathematicians approach mathematical tasks.

Participants' difficulties with the Sine Task

We predicted that constructing the required proof would not be trivial. The data displayed in Table 2 suggests that this expectation was correct.

The task was not straightforward for the participants as a whole. Seven participants spent more than 10 minutes working on this task, with two spending more than 30 minutes. While working on the proof construction, seven participants made assertions that were either false or true but irrelevant to the proof that they produced. Only KT's proof construction (section 3.4) proceeded directly.

While false starts and missteps are common in mathematical problem solving, even among mathematicians, we further note that only four participants produced a completely correct proof, even though all participants were invited to revise their initial proofs. Two participants had what we judged to be minor errors in their proofs, such as TL's proof in section 3.2. FR (section 3.4) produced an invalid proof, and BT reached an impasse after eleven minutes, producing the proof only after asking for a hint.

Finally, we note that there were cases in which these mathematicians had the essential idea for how their proof would proceed, yet still had difficulty in finalising the proof. For instance, section 3.2.2 illustrates how TL had difficulty proving functions cannot be injective on intervals with local maxima, despite the fact that the claim was "sort of obviously" true to him and that he had a diagram which clearly illustrated the main idea.

Purposes of diagrams in participants' proof constructions

From our analysis, we identified four purposes that participants' diagrams served. These purposes are summarised in Table 3. Note that we do not claim that all participants, or all mathematicians in general, would use diagrams for each of these purposes. Indeed, only KZ used diagrams for all four purposes in his proof attempt. Rather, we illustrate the ways that individual participants used diagrams.

Noticing properties and generating conjectures

A common way in which participants used diagrams was to note properties that were, or might be, true about the sine function. Specifically, we coded a statement in this category if it was coded as an assertion, the assertion was not deduced from any previous statements, and the source of the assertion was a diagram.

This purpose for using diagrams was illustrated in section 3.1.1., where KZ inspected the sine graph and noticed the property $\sin(\pi/2+\theta)=\sin(\pi/2-\theta)$. In section 3.3.1., FR, after using the graph of the sine function to reject that the sine function had period π, noticed the property that sine had period 2π (although this property was not used in his proof).

Finally, we note that even when participants noticed true properties about the sine function, they still sometimes felt the need to verify these properties deductively.

Table 3. A summary of the mathematicians' purposes of using diagrams.

Purpose served by diagram	Total number of instances	Participants who used the diagram for this purpose
Noticing properties and generating conjectures (NP)	11	TL(3), CY(2), PV(3), FR(1), IR(1), KZ(1)
Representing/ instantiating an idea or assertion (RI)	6	TL(3), KZ(2), BT(1)
Estimating the truth of an assertion (ET)	9	PV(4), FR(1), IR(2), KZ(2)
Suggesting a proof approach (SPA)	6	TL(4), KZ(1), BT(1)

For instance, much of KZ's proof attempt in section 3.1 consists of KZ trying to verify a generalised version of $\sin(\pi/2+\theta)=\sin(\pi/2-\theta)$.

Estimating the truth of an assertion

When participants were unsure if an assertion that they made was true, they would sometimes attempt to verify the assertion with a diagram. If the diagram appeared to affirm the assertion, this would increase the participant's confidence that the assertion was correct. Specifically, we coded a statement in this category if (a) the statement was coded as an evaluation, (b) the source of the evaluation was a diagram, and (c) the result of this evaluation appeared to increase or decrease the participant's confidence that a previously made assertion was true.

In some cases, if the diagram seemed inconsistent with a prior assertion, the participant would reject that assertion as false. This occurred in section 3.3.1, when FR used the sine graph to recognise that his initial claim that $\sin(x)=\sin(x+\pi)$ was false.

In other cases, participants were unsure if an identity that they recalled was correct, and turned to a diagram for confirmation. This is illustrated in 3.1.2, where KZ uses a diagram to check if the identity $\sin(\theta)=\cos(\pi/2-\theta)$ was correct by inspecting a labeled right-angled triangle. Later, in section 3.1.4, he verified the validity of the same equation by looking at a graph of the sine and cosine functions.

Suggesting a proof approach

Participants sometimes used a diagram to suggest an overarching plan of attack for proving the desired statement. We coded a statement in this category if the statement was coded as a proof approach, and the source of the statement was a diagram.

Perhaps the strongest illustration of this was in TL's work in section 3.2.1. Here, the graph in Figure 2 played a key role in his development of a sketch of the proof. In this episode, TL gestures to the graph while formulating his plan to translate an arbitrary interval back to a particular region on the graph. Later on in his interview, TL focuses on proving that a continuous function cannot be injective on a neighborhood of a local maximum (section 3.2.2.), and draws the diagrams in Figure 2 as aids to suggest ways to prove this claim. Finally, we note that BT used a unit circle to make a systematic search for points (angles), the sines of which were the same. However, he abandoned this approach after several minutes of failing to make progress.

Instantiating or representing an idea or assertion in a diagram

We coded an instance in this category if participants made an assertion or a proof approach, and either created a new diagram or modified an existing diagram to represent these ideas visually. Note that this purpose differs from noticing properties and generating conjectures – the former involves participants deducing an assertion from a diagram, while the latter has participants altering a diagram based on an assertion. A possible motive for doing this was that a visual representation of a mathematical assertion or idea might suggest inferences or proof approaches that might be useful.

To illustrate, when KZ asserted that every interval of length greater than length π must contain an element of the form $k\pi/2$ for some odd integer k, he drew a real number line and an arbitrary interval of length greater than π to illustrate his claim (see section 3.1.3). Similarly, in section 3.2.1, TL is continually modifying Figure 2 to represent the ideas he is stating aloud. This modification went hand in hand with the development and articulation of a proof approach.

Discussion

As we only analysed the behaviour of eight mathematicians proving a single claim, we cannot make general claims about mathematicians' behaviour with diagrams and proofs. Indeed, our data suggest that there might not be a single way to describe mathematicians' interactions with diagrams, as the participants did not exhibit strong heterogeneity in their behaviour, either with the extent of their diagram use or with the purposes for which they used diagrams.

From our perspective, participants had a surprising amount of difficulty in their proof construction. Indeed, participants still had trouble after they articulated what might be called the crux of the proof. KZ had significant difficulties finalising his proof, due to difficulties in justifying an important trigonometric identity. TL had trouble proving an important relationship between injectivity on intervals and local maxima, despite having a diagram that made the relationship "sort of obviously" true.

We contend that these episodes illustrate an important point. Some researchers have claimed that the essence of a proof is contained in the intuitive ideas used to develop the formal proof (e.g. Hanna 1991). While Hanna acknowledges that having a logically and formally correct argument is important, she refers to the rigour as a "hygiene factor", implying that translating an intuitive idea to a formal proof is a relatively uninteresting and routine part of the proving process for mathematicians. Although this is probably true, the difficulties experienced by the participants in this study remind us that the process is still non-trivial. If some mathematicians experience such difficulties, students are likely to as well. Consequently, although the generation of informal visual arguments to serve as a basis for formal proofs is a laudable goal in the advanced mathematics classroom, care must still be taken in describing the complex process of translating these arguments into a proof (e.g. Alcock 2010; Duval 2007).

Our analysis revealed a total of four ways in which these mathematicians used diagrams: noticing properties and generating conjectures, estimating the truth of an assertion, suggesting a proof approach, and instantiating an idea or assertion. Previous studies on students' proof constructions using diagrams suggest that undergraduates use diagrams primarily to try to understand a mathematical claim, and sometimes to generate an explanation of why a theorem is true (see Alcock and Weber 2010a, 2010b). However, these students generally did not use diagrams for the same reasons that some of our participants did, such as verifying that their logical deductions were true, uncovering errors in their logic, or identifying false mathematical assertions. Further, in mathematics lectures that we have observed, these aspects of proof construction were often absent (e.g. Weber 2004). We therefore suggest that the specific ways in which mathematicians used diagrams in this study be highlighted for students during instruction. Of course, direct instruction is not guaranteed to improve students' ability to use diagrams to construct proofs, but we argue that diagrammatic reasoning deserves explicit attention in the undergraduate mathematics classroom. Finally, as Alcock (2010) notes, more research is needed on instruction that could help students use diagrams effectively. Consequently, this study can be viewed as the basis for the tentative and untested pedagogical recommendation that mathematical lectures model the different ways that some mathematicians use diagrams, and we suggest future research on how we can improve students' usage and appreciation of diagrams in proof construction.

Acknowledgements

We are especially grateful to the mathematicians who participated in this study, and who so graciously allowed us to benefit from the depth of their mathematical experience – we found the thoughts of each mathematician fascinating and insightful. We appreciate their generosity in opening their work and reflections on the study tasks to analysis. We thank the editor and the anonymous reviewers for helpful comments. This work was supported by grants from the National Science Foundation (#EHR-1008317, #DUE-081736, and #DRL-0643734).

References

Alcock, L. 2010. Interactions between teaching and research: Developing pedagogical content knowledge for real analysis. In *Learning through teaching mathematics*, ed. R. Leikin and R. Zazkis, 227–46. Springer: Dordrecht.

Alcock, L.J., and A.P. Simpson. 2004. Convergence of sequences and series: Interactions between visual reasoning and the learner's beliefs about their own role. *Educational Studies in Mathematics* 57, no. 1: 1–32.

Alcock, L.J., and K. Weber. 2005. Proof validation in real analysis: Inferring and checking warrants. *Journal of Mathematical Behavior* 24: 125–34.

Alcock, L.J., and K. Weber. 2010a. Undergraduates' example use in proof production: Purposes and effectiveness. *Investigations in Mathematical Learning* 31, no. 1–22.

Alcock, L.J., and K. Weber. 2010b. Referential and syntactic approaches to proving: Case studies from a transition-to-proof course. *Research in Collegiate Mathematics Education* 7: 101–23.

Boero, P. 2007. *Theorems in schools: From history, epistemology, and cognition to classroom practice*. Rotterdam: Sense Publishers.

Burton, L. 2004. *Mathematicians as enquirers: Learning about learning mathematics*. Berlin: Springer.

Dreyfus, T. 1991. On the status of visual reasoning in mathematics and mathematics education. In *Proceedings of the 15th Conference of the International Group for the Psychology of Mathematics Education*, ed. F. Furinghetti, Vol. I, 33–48. Assisi, Italy: University of Genova.

Duval, R. 2007. Cognitive functioning and understanding the mathematical processes of proof. In *Theorems in schools: From history, epistemology, and cognition to classroom practice*, ed. P. Boero, 137–161. Rotterdam: Sense Publishers.

Gibson, D. 1998. Students' use of diagrams to develop proofs in an introductory real analysis. *Research in Collegiate Mathematics Education* 2: 284–307.

Hadamard, J. 1945. *The Psychology of invention in the mathematical field*. Princeton, NJ: Princeton University Press.

Hanna, G. 1991. Mathematical proof. In *Advanced mathematical thinking*, ed. D. Tall, 54–61. Dordrecht: Kluwer.

Harel, G., and L. Sowder. 1998. Students proof schemes. *Research in collegiate mathematics education* 3: 234–82.

Hart, E. 1994. A conceptual analysis of the proof writing performance of expert and novice students in elementary group theory. In *Research issues in mathematics learning: Preliminary analyses and results*, ed. J. Kaput and E. Dubinsky, 49–62. Washington, DC: Mathematical Association of America.
Hersh, R. 1993. Proving is convincing and explaining. *Educational Studies in Mathematics* 24, no. 4: 389–99.
Malone, J., G. Douglas, B. Kissane, and R. Mortlock. 1980. Measuring problem solving. In *NCTM 1980 yearbook: Problem solving in school mathematics*, ed. S. Kruklik, 204–215. Washington, DC: NCTM.
Moore, R.C. 1994. Making the transition to formal proof. *Educational Studies in Mathematics* 27: 249–66.
NCTM. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
Nelsen, R.B. 1993. *Proofs without words*. MAA: Washington, DC.
Nelsen, R.B. 2000. *Proofs without words II*. MAA: Washington, DC.
Piez, C.M., and M.H. Voxman. 1997. Multiple representations: Using different perspectives to form a clearer picture. *The Mathematics Teacher* 90: 164–6.
Pólya, G. 1957. *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.
Presmeg, N. 1986. Visualization and mathematical giftedness. *Educational Studies in Mathematics* 17: 297–311.
Raman, M. 2003. Key ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics* 52: 319–25.
Recio, A.M., and J.D. Godino. 2001. Institutional and personal meanings of proof. *Educational Studies in Mathematics* 48, no. 1: 83–99.
Schoenfeld, A. 1985. *Mathematical problem solving*. Orlando, FL: Academic Press.
Strauss, A., and J. Corbin. 1990. *Basics of qualitative research: Grounded theory procedures and techniques*. London: Sage.
Stylianou, D.A. 2002. On the interaction of visualization and analysis: The negotiation of a visual representation in problem solving. *Journal of Mathematical Behavior* 21, no. 3: 303–17.
Stylianou, D.A., and E.A. Silver. 2004. The role of visual representations in advanced mathematical problem solving: An examination of expert-novice similarities and differences. *Journal of Mathematical Thinking and Learning* 6, no. 4: 353–87.
Thurston, W.P. 1994. On proof and progress in mathematics. *Bulletin of the American Mathematical Society* 30: 161–77.
Weber, K. 2001. Student difficulties in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics* 48, no. 1: 101–19.
Weber, K. 2004. Traditional instruction in advanced mathematics courses: A case study of one professor's lectures and proofs in an introductory real analysis course. *Journal of Mathematical Behavior* 23, no. 2: 115–33.
Weber, K. 2006. Investigating and teaching the thought processes used to construct proofs. *Research in Collegiate Mathematics Education* 6: 197–232.
Weber, K., and L.J. Alcock. 2004. Semantic and syntactic proof productions. *Educational Studies in Mathematics* 56, no. 3: 209–34.

Appendix

In what follows, we give two correct proofs from our participants PV and KT. Both proofs were lightly edited for the purposes of readability and brevity, but the central ideas employed in the proofs were unchanged. Recall that the Sine Task asks participants to show the claim that *restrictions of the sine function to intervals of length greater than* π *cannot be injective.*

Proof 1 (PV) – Using reflectional symmetry of the sine function:

Every interval of length greater than π contains a point x_0 of the form $\frac{\pi}{2} + \pi k$ as well as a neighbourhood I containing x_0, where k is an integer. Now, sine is symmetric about $\frac{\pi}{2} + \pi k$ in the sense that $\sin(\frac{\pi}{2} + \pi k + x) = \sin(\frac{\pi}{2} + \pi k - x)$ for all real x and integer k, as can be checked using the sine angle addition formula. Choose ε such that $x_0 - \varepsilon$ and $x_0 + \varepsilon$ are elements of I. Then $\sin(x_0 + \varepsilon) = \sin(\frac{\pi}{2} + \pi k + \varepsilon) = \sin(\frac{\pi}{2} + \pi k - \varepsilon) = \sin(x_0 - \varepsilon)$, so sine is not injective on I, and hence not injective on the original interval.

Proof 2 (KT) – Using the Intermediate Value Theorem:

Every interval of length greater than π contains a point x_0 of the form $\frac{\pi}{2} + k\pi$ in its interior. If $f(x) = \sin(x)$, then $f'(x) = \cos(x)$ and $f''(x) = -\sin(x)$. Note that $f'(x_0) = 0$ and $f''(x_0) \neq 0$. Hence f attains a relative maximum or minimum at x_0. Assume f attains a relative maximum at x_0, and I is a neighbourhood of x_0 on which x_0 is a maximum. (an argument similar to the one below may be applied to the case where x_0 is a relative minimum). Let a, b be in I such that $a < x_0 < b$. Let $y_0 = \max(f(a), f(b))$. Then, since $f(x_0)$ is a maximum on I, we have $y_0 \leq f(x_0)$. Also, $f(a) \leq y_0$ and $f(b) \leq y_0$ by construction. Since f is continuous, we may apply the Intermediate Value Theorem: there exist a' in $[a, x_0]$ and b' in $[x_0, b]$ such that $f(a') = y_0 = f(b')$. Hence, $f(x) = \sin(x)$ is not injective on I, and hence not injective on the original interval.